Journal of Forensic and Legal Medicine 20 (2013) 112–116

Contents lists available at SciVerse ScienceDirect

Journal of Forensic and Legal Medicine

journal homepage: www.elsevier.com/locate/jflm

Original communication

# Evaluation of the allele-sharing approach, known as the IBS method, in kinship analysis

Tomonori Tamura MPharm, Forensic Technologist [a], Motoki Osawa MD, PhD, Professor [b,*], Ryosuke Kimura PhD, Associate Professor [c], Yoshihiko Inaoka PhD, Assistant Professor [b], Sachi Tanaka MPAS, Forensic Technologist [a], Fumiko Satoh MD, PhD, Associate Professor [b], Itaru Sato PhD, Division Director [a]

[a] Scientific Crime Laboratory, Kanagawa Prefectural Police, Yamashitacho 155, Naka-ku, Yokohama 231-0023, Japan
[b] Department of Forensic Medicine, Tokai University School of Medicine, Shimokasuya 143, Isehara, Kanagawa 259-1193, Japan
[c] Transdisciplinary Research Organization for Subtropics and Island Studies, University of the Ryukyus, Nishihara, Okinawa 903-0215, Japan

ARTICLE INFO

*Article history:*
Received 6 August 2010
Received in revised form
10 February 2011
Accepted 30 May 2012
Available online 26 June 2012

*Keywords:*
Identical by state
Short tandem repeat
Combined sibship index
Genetic relatedness
Heterozygosity

ABSTRACT

To infer relatedness from genetic data based on short tandem repeats, the exact method, in which shared allele frequencies are applied to relevant equations, has been conventionally used. An alternative approach is the IBS method that is based on the number of shared alleles between individuals. In the present study, the performance of the IBS method in pairwise kinship analysis was compared with the exact method using simulated data of 10,000 genotype pairs for 15 loci in the ABI Identifiler system. The likelihood ratio in allele-sharing of zero, one and two was calculated from joint probabilities based on allele frequencies of the Japanese population. Whereas the IBS method generally produced lower values of combined indices, smaller deviations of the distributions were evident. The threshold for identification of full siblings relative to non-relatives was comparable with that of the exact method, indicating that both inference powers were almost identical. The likelihood ratio in the IBS method depends on the heterozygosity at a locus, and heterozygosities of the 15 loci were consistent across various population groups, particularly in East Asians. The convenience of fixed LR values in the IBS method is beneficial for cases with uncertain allele frequencies and rare alleles.



## 1. Introduction

Multiplex analysis of short tandem repeats (STRs) has been extensively employed for the last decade for identification of missing individuals and unknown human remains, in addition to physical examinations including fingerprints, dental work and other characteristics. Direct comparison with personal remains provides the most meaningful conclusion based on type matching or two-allele sharing in all examined loci. However, in cases with an absence of personal remains, it may be necessary to prove a close biological relationship with a living relative.

The common method to infer a biological relationship from pairwise genotype data in loci is based on population frequencies of the observed alleles that were shared by the pair of individuals, and on probability equations for genotype combinations.[1,2] The likelihood ratio (LR) expresses the probability ratio of putative relatives to non-relatives. However, we have encountered a couple of difficulties in these cases. For instance, ethnic reversion may be unclear for foreign individuals, or the population frequencies of alleles may be unknown.

The allele sharing approach refers simply to the number of shared alleles at a locus between two individuals; these are also referred to as identical-by-state (IBS) alleles.[3] The LR in this alternative approach is based on probabilities of the shared allele numbers of zero, one and two, denoted as $z_i$ and calculated in advance from population data of allele frequencies in an interest group at Hardy–Weinberg equilibrium. Presciuttini et al.[4] successfully developed this IBS method, which was originally proposed by Chakraborty and Jin,[5] for inference of a pairwise relationship using Caucasian STR data. It was shown that the LR values were functionally dependent on the locus heterozygosity ($H$), rather than on allele frequencies. However, to our knowledge, no further evaluation has been attempted, except for studies on incidences of allele sharing in various relationships.[6,7]

* Corresponding author. Tel.: +81 463 93 1121; fax: +81 463 92 0284.
*E-mail address:* osawa@is.icc.u-tokai.ac.jp (M. Osawa).

1752-928X/$ – see front matter © 2012 Elsevier Ltd and Faculty of Forensic and Legal Medicine. All rights reserved.
http://dx.doi.org/10.1016/j.jflm.2012.05.005

A multiplex STR system, the AmpFℓSTR Identifiler kit, is the national standard for crime scene investigations in Japan, and is used worldwide. The system consists of 15 autosomal STR loci (13 CODIS loci, D2S1338 and D19S433) and the sex-identifying locus Amelogenin. The multiplex analysis effectively solves a variety of genetic problems, including personal identification and paternity test, through assessment based on indices combined with the unlinked 15 loci. In indirect comparisons with a known family member, the strongest conclusion can be achieved for a parent–child relationship because obligatory single or two-allele sharing usually produces a value of over 1000 for the combined parentage index (CPI) that is obtained by multiplying the LR values. In the full-sibling relationship, a lack of shared alleles at a locus does not exclude two persons from being related,[1] and incomplete separation of the distributions of the combined sibship index (CSI) between true siblings and non-relatives usually occurs. Pu and Linacre[7] demonstrated the improved determination of sibship using the combination of CSI and two-allele-sharing loci. Giroti et al.[8] showed that the CSI value from 0.067 to 10.3 was in the gray zone. Moreover, CSI thresholds to demarcate potential non-relatives were chosen as cut-off points of 1 by Reid et al.[6] and 3 by Tzeng et al.[9] in the conventional exact method. In the present study, the IBS method for kinship analysis was evaluated for the potentiality of the more common usage. A comparison with the exact method is performed using simulated and observed data from STR profiles obtained with the ABI Identifiler system.

## 2. Materials and methods

### 2.1. Subjects and genotyping

DNA extraction was carried out with BioRobot EZ1 (Qiagen, Hilden, Germany) using an EZ1 DNA Investigator kit, according to the manufacturer's instructions. Extracted DNA was added to the reaction mixture of the AmpFℓSTR Identifiler kit (Applied Biosystems, Foster City, CA) in a tube, and amplified using a GeneAmp PCR System 9700. Amplicons were separated and detected using a 3130xl Genetic Analyzer (Applied Biosystems) with reference to Liz 500 size standards. Genotyping was automatically performed using Genemapper ID ver. 3.2.1 (Applied Biosystems). A total of 478 Japanese subjects comprising 135 parent–child pairs and 104 full-sibling pairs were genotyped in two departments of forensic science. Another 546 non-relative pairs were constructed by random combination of selected profiles. The study was approved by the ethics committee of Tokai University School of Medicine.

### 2.2. LR in the IBS method

The joint probability implies that a pair of persons would have genotypes $G_1$ and $G_2$. In a multi-allelic locus consisting of more than four alleles, seven distinct patterns of allele sharing can appear between a pair: AA–AA, AA–AB, AA–BB, AB–AB, AA–BC, AB–AC, and AB–CD, where A, B, C, and D denotes different alleles in a locus. The expected joint probability of each combination was calculated by equations for the three relationships of parent–child, full siblings and non-relatives,[10] to which allele frequencies for the Japanese population, reported by Yoshida et al.,[11] were applied. Then, the allele sharing probability ($z_i$) was obtained by summing up the joint probabilities: three $G_1$–$G_2$ combinations of AA–BB, AA–BC and AB–CD for $z_0$, two of AA–AB and AB–AC for $z_1$, and another two of AA–AA and AB–AB for $z_2$.

Two alternative hypotheses ($H_S$, the two persons are full siblings; $H_0$, the two persons are non-relatives) were considered with $LR_i$ for a single locus given by $P(z_i|H_S)/P(z_i|H_0)$. As long as there is no genetic linkage among STR loci, the overall indices can be determined by multiplying all single locus $LR_i$ that corresponds to PI in a parentage test and SI in a sibship test. For comparison of the results from the exact and IBS methods, the indices were converted to logarithms (base 10).

To reveal the distribution of the cumulative locus number of IBS alleles in the set of 15 STR loci between two persons ($N_i$), all possible $z_i$ combinations of the 15 loci were listed. For example, when the number of zero-allele sharing loci is equal to 1, $N_0$ is given by the following equation.

$$N_0 = \sum\{z_{0j} \times \Pi(1 - z_{0k})\} \quad (j \neq k)$$

Statistical comparison of calculated data with experimental data was performed by calculating the Pearson correlation coefficient ($r$).

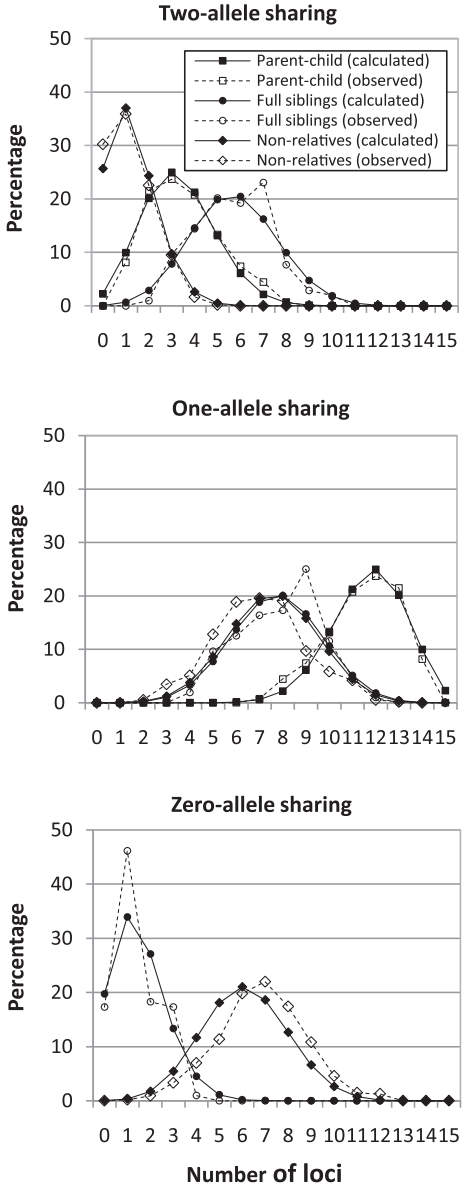


**Fig. 1.** Distribution of allele sharing instances of two (top), one (middle) and zero (bottom) in parent–child, full siblings and non-related pairs. The solid and broken lines represent the percentage calculated from the reported allele frequencies[11] and that in the present study, respectively.

### 2.3. Sensitivity and specificity

To examine the distribution of combined indices, and to determine the cut-off point in the sibship test, we performed a simulation study. In the simulation, assuming parent–child, siblings and non-relatives pairwise genotypes were randomly constructed according to the reported allele frequencies[11] and identical-by-descent (IBD) probabilities.[12] CSI values were calculated for the simulated pairs according to the exact and IBS methods, and designated as $CSI_{exact}$ and $CSI_{IBS}$, respectively. The sensitivity and specificity under variable cut-off values of the CSIs were then examined for the simulated data for both methods. Based on Gaytmenn et al.,[13] sensitivity and specificity were defined as the proportion of true siblings with CSI values greater than the threshold, and the proportion of non-relative pairs with CSI values less than the threshold, respectively. The results are shown as receiver operating characteristic (ROC) plots. The positive predictive value (PPV) and the negative predictive value (NPV) mean the proportion of subjects correctly identified as siblings, and that of subjects correctly identified as non-siblings, respectively.[7,13]

### 2.4. Comparison of heterozygosity

To compare the heterozygosity ($H$) among ethnic groups, expected $H$ values were obtained from data for Korean,[14] Taiwanese,[15] west Chinese (Han),[16] east Chinese,[17] Minnesota (Caucasian, native-American and Hispanic),[18] Greek,[19] Brazilian[20] and Ugandan[21] populations.

## 3. Results and discussion

### 3.1. LR in the IBS method

In order to develop the IBS method, the calculated $z_i$ and $LR_i$ at a single locus were obtained for parent–child, full siblings and non-relatives from the joint probability of seven possible combinations using Japanese population data, as summarized in the Supplementary Table. In the parent–child hypothesis, $z_1$ and $z_2$ were consistent with $H$ and $(1 - H)$, respectively, of the relevant locus. As shown by Presciuttini et al.,[4] $LR_i$ values in full siblings and non-relatives also depend on $H$ of a locus, with $H$ values ranging from 0.647 (TPOX) to 0.870 (D2S1338). Indeed, the values were well predicted by polynomial functions of $H$ inductively (data not shown). In the IBS method, $LR_i$ values were defined as certain fixed values regardless of the frequencies of the shared alleles. LRs $< 1$ imply that the genetic evidence indicates that the putative family member is less biologically related. In the sibship test, one allele sharing is less attributable to CSI with the $LR_1$ values from 0.81 to 1.38.

The normal distribution of the cumulative number ($N_i$) of loci that have i IBS alleles is indicated in Fig. 1. The expected $N_i$ distributions from the reported allele frequencies[11] showed significant consistency with the present observed values obtained in the STR profiling. In use of the Identifiler system, no significant difference in

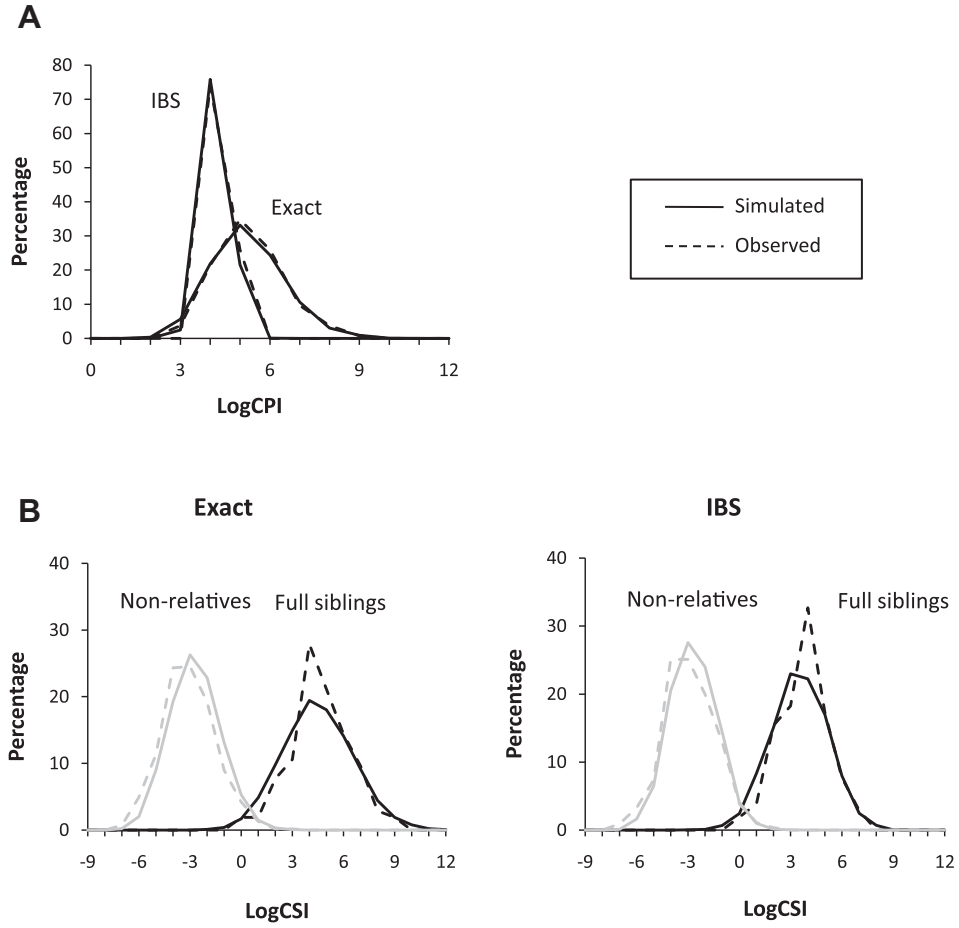


**Fig. 2.** Comparison of the distribution of combined indices between the two analytical methods. A, $CPI_{exact}$ and $CPI_{IBS}$ for true parent–child pairs, excluding non-relatives. B, $CSI_{exact}$ and $CSI_{IBS}$ of true siblings and non-relatives in the hypothesis for full siblings. Simulated and observed results are indicated with solid and broken lines, respectively. Lines for non-relatives are shown in gray. The $X$-axis is shown on a logarithmic (base 10) scale.

**Table 1**
Comparison of sensitivity, specificity, PPV and NPV in the two approaches.

| CSI threshold | Exact | | | | IBS | | | |
|---|---|---|---|---|---|---|---|---|
| | Sensitivity | Specificity | PPV[a] | NPV[b] | Sensitivity | Specificity | PPV | NPV |
| 0.01 | 0.999 | 0.797 | 0.831 | 0.998 | 0.999 | 0.803 | 0.835 | 0.998 |
| 0.03 | 0.998 | 0.872 | 0.886 | 0.998 | 0.997 | 0.894 | 0.904 | 0.996 |
| 0.1 | 0.995 | 0.928 | 0.932 | 0.994 | 0.992 | 0.948 | 0.950 | 0.992 |
| 0.3 | 0.989 | 0.961 | 0.962 | 0.989 | 0.985 | 0.972 | 0.973 | 0.984 |
| 1 | 0.978 | 0.981 | 0.981 | 0.978 | 0.968 | 0.988 | 0.988 | 0.968 |
| 3 | 0.959 | 0.990 | 0.990 | 0.960 | 0.937 | 0.995 | 0.995 | 0.941 |
| 10 | 0.930 | 0.996 | 0.996 | 0.934 | 0.884 | 0.998 | 0.998 | 0.896 |
| 33 | 0.885 | 0.998 | 0.998 | 0.897 | 0.819 | 1.000 | 1.000 | 0.847 |
| 100 | 0.831 | 0.999 | 0.999 | 0.856 | 0.734 | 1.000 | 1.000 | 0.790 |
| 333 | 0.761 | 1.000 | 1.000 | 0.807 | 0.612 | 1.000 | 1.000 | 0.721 |
| 1000 | 0.682 | 1.000 | 1.000 | 0.759 | 0.531 | 1.000 | 1.000 | 0.681 |

[a] PPV: positive predictive value.
[b] NPV: negative predictive value.

the incidence of one-allele sharing has been observed between full siblings and non-relatives.[6,7] Therefore, the numbers of zero- and two-allele sharing mostly determine the CSIs, and the incidental extremity affects the variance. For instance, the $N_0$ incidences at 0, 1, 2 and ≥3 in true siblings are expected to be 0.20, 0.34, 0.27 and 0.19, respectively. In contrast, the $N_0$ incidence at ≥3 in non-relatives is expected to be 0.98.

### 3.2. Combined indices in the parentage and sibship tests

In the previous study by Presciuttini et al.,[4] the two analytical methods were compared using the data from the limited numbers such as 80 sib pairs. To extensively evaluate the combined indices of $LR_i$ in the 15 core STR loci, pairs based on 10,000 simulations were constructed for parent–child, full siblings and non-relatives. Distributions of the combined indices in the parentage and sibship tests are shown in Fig. 2. Using the exact method, the logarithmic values of $CPI_{exact}$ and $CSI_{exact}$ were distributed with a mean (±S.D.) of 4.74 ± 1.20 and 3.99 ± 2.07 for true parent–child and sibling pairs, respectively, consistent with data reported by Tamaki et al.[22] The mean $CPI_{IBS}$ and $CSI_{IBS}$ values in the IBS method were 3.69 ± 0.41 and 3.01 ± 1.65, respectively. The observed data confirmed the analytic results derived from simulation. It is of note that the IBS method gave smaller deviations than the conventional exact approach, in particular for the parentage test.

The Pearson correlation coefficient ($r$) between combined indices of the two methods was 0.219 for parent–child pairs and 0.875 for full sibling pairs. Presciuttini et al.[4] found values of $r$ of 0.789 and 0.892, respectively, for these comparisons. The reason for the large difference in the correlation coefficients for the parentage test is unclear.

In the parentage test, 2.5% of the simulated parent–child pairs had $CPI_{IBS} < 1000$, in contrast to 6.0% with $CPI_{exact} < 1000$ in the exact method using allele frequencies. Complete separation between parent–child and non-relatives was evident in both analytical methods. In the sibship test, $CSI_{IBS}$ for true siblings was <100 in 26.6% and <1000 in 49.6% of simulated cases in the IBS method, whereas $CSI_{exact}$ was <100 and <1000 in 16.9% and 31.8% of cases, respectively. In common, combined indices > 1000 provide very strong evidence in favor of the hypothesized relationships in the exact method.[13] Therefore, the IBS method could seem to have less inference power in the sibling test, but the certainty threshold has to be ensured further.

### 3.3. Sensitivity and specificity of CSI

To obtain the cut-off value for $CSI_{IBS}$, the sensitivity, specificity, PPV and NPV were examined in the simulated pairs (Table 1). The ROC curves for the exact and IBS methods are shown in Fig. 3. In the IBS method, the negative predictive value and accuracy were optimized when 1 in logarithm was adopted as the cut-off value.[13] Full-sibling pairs were correctly judged with an incidence of 0.968 at that threshold, and non-relative pairs were correctly rejected with an incidence of 0.988. The exact method gave rise to more false positive results at any cut-off values. It is notable that the specificities of both methods are comparable at the same sensitivity levels.

Reid et al.[23] indicated that finding true siblings in a large forensic database was difficult using the 13 CODIS core loci with an available analytic algorithm. It is likely that use of insufficient loci gives rise to incomplete separation. A recent study by Nothnagel et al.[24] demonstrated that analysis of a total of 34 STR loci can overcome this difficulty in a sib genetic test, in which consideration of the genetic linkage of the loci appear to be required.

### 3.4. Comparison of allele frequencies in various ethnic populations

Presciuttini et al.[4] compared $H$ values in the 13 core STR loci among a variety of Caucasian population groups. A comparison of the reported $H$ values among East Asians and other groups is shown in Fig. 4. Despite small deviations, no critical differences in $H$ values were evident among the close Chinese, Korean, and Japanese population groups. This indicates that the IBS method using standardized LR values is applicable for East Asian populations. In addition, Pu and Linacre[7] demonstrated no significant difference in distributions of shared allele instances among three populations of Han Chinese,

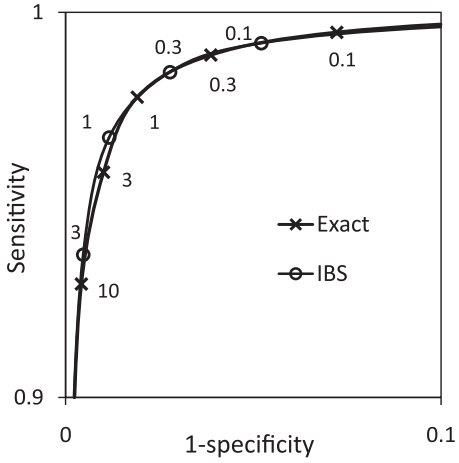


**Fig. 3.** Receiver operating characteristic (ROC) plots for the two methods at several cut-off values. Numbers represent the applied cut-off points for $CSI_{IBS}$ and $CSI_{exact}$. $X$- and $Y$-axes indicate the false-positive rate and sensitivity, respectively.

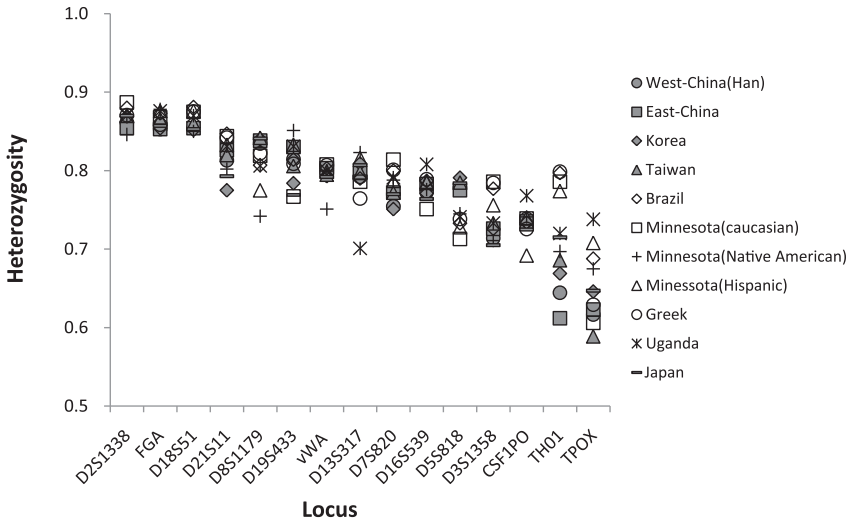


**Fig. 4.** Comparison of expected *H* values at the 15 STR loci in the Identifiler system among 11 population groups.[11,14–21]

Caucasians and African Americans, suggesting that universal LR values for some STR loci are potentially developed in the IBS method.

In conclusion, the present study focused on use of LRs and combined indices as a standard statistical procedure in a biological relationship test, excluding conditional probabilities based on the Bayes's theorem. To develop the IBS method for pairwise relatedness analysis, we constructed LR values for the 15 core loci of the Identifiler system, based on allele frequencies for the Japanese population. The IBS method constrains the distribution of combined kinship indices, whereas the exact method occasionally produces extreme $LR_i$ values generated by rare alleles and variants. Moreover, the optimized cut-off point of $CSI_{IBS}$ was obtained as 1, indicating that the inference power of the IBS method is comparable with that of the conventional approach. Preliminary application of the IBS method might be reasonable in cases in which exact allele frequencies are unavailable for subjects, due to the lower variability of *H* across population groups. As the alternative of the exact method, these may be the main advantages of the IBS approach.

*Conflict of interest*

None declared.

*Funding*

This work was supported by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

*Ethical approval*

The study was approved by the ethics committee of Tokai University School of Medicine.

## Appendix A. Supplementary data

Supplementary data related to this article can be found online at http://dx.doi.org/10.1016/j.jflm.2012.05.005.

## References

1. Wenk RE, Traver M, Chiafari FA. Determination of sibship in any two persons. *Transfusion* 1996;**36**:259–62.
2. Ayres KL. Relatedness testing in subdivided populations. *Forensic Sci Int* 2000;**114**:107–15.
3. Aoki Y, Nakayama Y, Saigusa K, Nata M, Hashiyada M. Comparison of the likelihood ration and identity-by-state scoring methods for analyzing sip-pair test cases: a study using computer simulation. *Tohoku J Exp Med* 2001;**194**:241–50.
4. Presciuttini S, Toni C, Tempestini E, Verdiani S, Casarino L, Spinetti I, et al. Inferring relationship between pairs of individuals from locus heterozygosities. *BMC Genet* 2002;**3**:23.
5. Chakraborty R, Jin L. Determination of relatedness between individuals using DNA fingerprinting. *Hum Biol* 1993;**65**:875–95.
6. Reid TM, Wolf CA, Kraemer CM, Lee SC, Baird ML, Lee RF. Specificity of sibship determination using the ABI Identifiler multiplex system. *J Forensic Sci* 2004;**49**:1262–4.
7. Pu CE, Linacre A. Systematic evaluation of sensitivity and specificity of sibship determination by using 15 STR loci. *J Forensic Leg Med* 2008;**15**:329–34.
8. Giroti RI, Verma S, Singh K, Malik R, Talwar I. A grey zone approach for evaluation of 15 short tandem repeat loci in sibship analysis: a pilot study in Indian subjects. *J Forensic Leg Med* 2007;**14**:261–5.
9. Tzeng CH, Lyou JY, Chen YR, Hu HY, Lin JS, Wang SY, et al. Determination of sibship by PCR-amplified short tandem repeat analysis in Taiwan. *Transfusion* 2000;**40**:840–5.
10. Buckleton J, Triggs CM. Relatedness. In: Buckleton J, Triggs CM, Walsh SJ, editors. *Forensic DNA evidence interpretation*. Boca Raton: CRC Press; 2005. p. 123–48.
11. Yoshida K, Takahashi K, Kasai K. Allele frequencies of 15 loci using AmpFlSTR Identifiler Kit in Japanese population. *J Forensic Sci* 2005;**50**:718–9.
12. Thomson JA, Ayres KL, Pilotti V, Barrett MN, Walker JI, Debenham PG. Analysis of disputed single-parent/child and sibling relationships using 16 STR loci. *Int J Legal Med* 2001;**115**:128–34.
13. Gaytmenn R, Hildebrand DP, Sweet D, Pretty IA. Determination of the sensitivity and specificity of sibship calculations using AmpFl STR Profiler Plus. *Int J Legal Med* 2002;**116**:161–4.
14. Kim YL, Hwang JY, Kim YJ, Lee S, Chung NG, Goh HG, et al. Allele frequencies of 15 STR loci using AmpFlSTR Identifiler Kit in a Korean population. *Forensic Sci Int* 2003;**136**:92–5.
15. Wang CW, Chen DP, Chen CY, Lu SC, Sun CF. STR data for the AmpFlSTR SGM Plus and Profiler loci from Taiwan. *Forensic Sci Int* 2003;**138**:119–22.
16. Deng YJ, Yan JW, Yu XG, Li YZ, Mu HF, Huang YQ, et al. Genetic analysis of 15 STR loci in Chinese Han population from West China. *Genomics Proteomics Bioinformatics* 2007;**5**:66–9.
17. Li C, Li L, Zhao Z, Lin Y, Que T, Liu Y, et al. Genetic polymorphism of 17 STR loci for forensic use in Chinese population from Shanghai in East China. *Forensic Sci Int Genet* 2009;**3**:e117–8.
18. Gross AM, Budowle B. Minnesota population data on 15 STR loci using the Identifiler Kit. *J Forensic Sci* 2006;**51**:1410–3.
19. Sánchez-Diz P, Menounos PG, Carracedo A, Sitsa I. 16 STR data of a Greek population. *Forensic Sci Int Genet* 2008;**2**:e71–2.
20. de Assis Poiares L, de Sá Osorio P, Spanhol FA, Coltre SC, Rodenbusch R, Guesmão L, et al. Allele frequencies of 15 STRs in a representative sample of the Brazilian population. *Forensic Sci Int Genet* 2010;**4**:e61–3.
21. Gomes V, Sánchez-Diz P, Alves C, Gomes I, Amorim A, Carracedo A, et al. Population data defined by 15 autosomal STR loci in Karamoja population (Uganda) using AmpFlSTR Identifiler. *Forensic Sci Int Genet* 2009;**3**:e55–8.
22. Tamaki K, Kaszynski RH, Yuan QH, Yoshida K, Okuno T, Tsuruyama T. Likelihood evaluation using 15 common short tandem repeat loci: a practical and simulated approach to establishing personal identification via sibling/parental assessments. *Transfusion* 2009;**49**:578–84.
23. Reid TM, Baird ML, Reid JP, Lee SC, Lee RF. Use of sibling pairs to determine the familial searching efficiency of forensic databases. *Forensic Sci Int Genet* 2008;**2**:340–2.
24. Nothnagel M, Schmidtke J, Krawczak M. Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. *Int J Legal Med* 2010;**124**:205–15.